Linear Regression Analysis under Sets of Conjugate Priors

Gero Walter, Thomas Augustin, Annette Peters

Department of Statistics
Ludwig-Maximilians-University Munich

GSF – National Research Center for
Environment and Health, Neuherberg

July 17th, 2007

introducing Annette Peters

- Head of Research Unit 'Epidemiology of Air Pollution Health Effects' of GSF - Institute of Epidemiology
- research on health effects of fine and ultrafine particles (AIRGENE, EPA STAR) and alpha-particles (radon)

introducing Gero Walter

current PhD. student under the guidance of Thomas Augustin
Department of Statistics,
Ludwig-Maximilians-University (LMU) Munich
Research group on interval probabilities (T. Augustin, K. Weichselberger, A. Wallner, C. Strobl, R. Hable)

introducing Gero Walter

- 2007 Receiving of *Diplom* (equivalent to a Master's degree)
- 2006 Diploma thesis „Bayes-Regression mit Mengen von Prioris – Ein Beitrag zur Statistik unter komplexer Unsicherheit“
- 2005 Internship at GSF – National Research Center for Environment and Health,
  Research Unit "Epidemiology of Air Pollution Health Effects",
  head Dr. A. Peters (AIRGENE study group)
- '00 – '07 Student of Statistics at the the Department of Statistics, LMU
- '02 – '03 University of Palermo, Italy (Erasmus programme)

Research interests

- imprecise probability models for linear and generalized linear regression
- Robust Bayesian approaches
- modeling of prior-data conflict

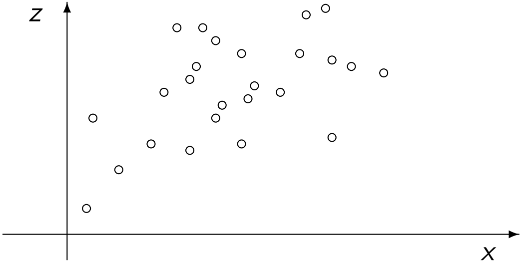
- Bayesian variable selection (LASSO, ...)
- Statistics in Epidemiology

Linear Regression Analysis under Sets of Conjugate Priors

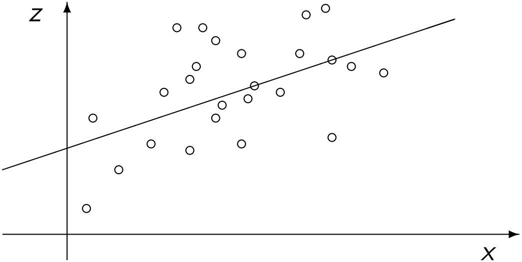
- Linear Regression Analysis
  - linear regression & generalizations
  - estimation methods
- Conjugate Priors
  - Bayesian estimation
  - LUCK-models
- Sets of Conjugate Priors
  - method of Quaeghebeur and de Cooman
    - direct application
    - generalizing the standard approach
  - the *imprecise normal regression model*
- Application and Concluding Remarks

Linear regression

Linear regression

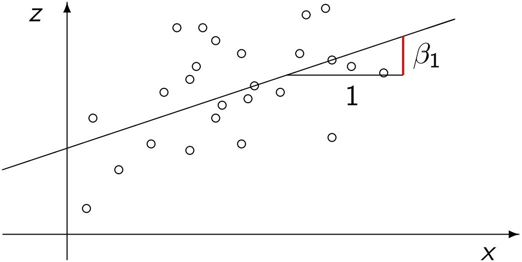


Linear regression



$$z_i = \beta_0 + x_i \cdot \beta_1 + \varepsilon_i$$

Linear regression



$$z_i = \beta_0 + x_i \cdot \beta_1 + \varepsilon_i$$

β_1 interpretable as increment on z if x increases one unit

Linear regression

$$z_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \,,$$

z_i	obs. i of response (dependent variable,...)
$(x_{i1}, \ldots, x_{ip}) =: x_i$	obs. i of regressors $j = 1, \ldots, p$ (independent variables,...)
$(\beta_1, \ldots, \beta_p) =: \beta$	regression coefficients
ε_i	stochastic error term

$$z = \mathbf{X}\beta + \varepsilon \,, \quad \mathbf{X} \in \mathbb{R}^{k \times p},\ \beta \in \mathbb{R}^p,\ z \in \mathbb{R}^k,\ \varepsilon \in \mathbb{R}^k \,;$$

$$\varepsilon_i \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2) \quad \Longrightarrow \quad \varepsilon \sim \mathrm{N}_k(\mathbf{0}, \sigma^2 \mathbf{I})) \quad (\sigma^2 \text{ known})$$

(one) categorial regressor x ▶ ANOVA

Linear regression

$$z_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \,,$$

z_i	obs. i of response (dependent variable,...)
$(x_{i1}, \ldots, x_{ip}) =: x_i$	obs. i of regressors $j = 1, \ldots, p$ (independent variables,...)
$(\beta_1, \ldots, \beta_p) =: \beta$	regression coefficients
ε_i	stochastic error term

$$z = \mathbf{X}\beta + \varepsilon \,, \quad \mathbf{X} \in \mathbb{R}^{k \times p},\ \beta \in \mathbb{R}^p,\ z \in \mathbb{R}^k,\ \varepsilon \in \mathbb{R}^k\,;$$

$$\varepsilon_i \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2) \quad \Longrightarrow \quad \varepsilon \sim \mathrm{N}_k(\mathbf{0}, \sigma^2 \mathbf{I})) \quad (\sigma^2 \text{ known})$$

(one) categorial regressor x ▶ ANOVA

Generalizations of linear regression

- non-normal response (GLM) (categorical ▶ classification)
- complex correlation structure of observations (e.g., repeated measurements, spatial,. . .)
- non-linear regressors (GAM)
- survival data analysis

one of the most important inference tools in all areas of application

Estimation of β

- Least Squares (LS) method: minimize $\sum_{i=1}^{k}(z_i - x_i\,\beta)^2$:

$$\hat{\beta}_{LS} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T} z\,.$$

- Maximum Likelihood (ML) method: maximize likelihood $z \mid \beta \sim \mathrm{N}_k(\mathbf{X}\beta, \sigma^2\mathbf{I})$ ($\mathbf{X}$ non-stochastic) ▶ $\hat{\beta}_{LS}$
- Bayesian method: choose prior on β, maximize posterior (take posterior expected value)
  - often: weak prior information ▶ "objective Bayesian" paradigm: take "noninformative" prior $\beta \propto$ const. ▶ $\hat{\beta}_{LS}$
  - conjugate prior: convenient choice, posterior of same parametrical class as prior ▶ choice of parameters?

Estimation of β

- Least Squares (LS) method: minimize $\sum_{i=1}^{k}(z_i - x_i\,\beta)^2$:

$$\hat{\beta}_{LS} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T} z\,.$$

- Maximum Likelihood (ML) method: maximize likelihood $z \,|\, \beta \sim \mathrm{N}_k(\mathbf{X}\beta, \sigma^2\mathbf{I})$ ($\mathbf{X}$ non-stochastic) ▶ $\hat{\beta}_{LS}$
- Bayesian method: choose prior on β, maximize posterior (take posterior expected value)
  - often: weak prior information ▶ "objective Bayesian" paradigm: take "noninformative" prior $\beta \propto$ const. ▶ $\hat{\beta}_{LS}$
  - conjugate prior: convenient choice, posterior of same parametrical class as prior
    ▶ choice of parameters?

> To overcome "dogma of ideal precision" (Walley), consider sets of priors, here: by sets of parameters

Conjugate Priors

- Linear Regression Analysis
  - linear regression & generalizations
  - estimation methods
- Conjugate Priors
  - Bayesian estimation
  - LUCK-models
- Sets of Conjugate Priors
  - method of Quaeghebeur and de Cooman
    - direct application
    - generalizing the standard approach
  - the *imprecise normal regression model*
- Application and Concluding Remarks

Bayesian estimation with conjugate priors

$$p(\vartheta \mid w) \propto f(w \mid \vartheta) \cdot p(\vartheta) \tag{1}$$

we distinguish certain standard situations (called *models with 'Linearly Updated Conjugate prior Knowledge'* (LUCK) here) of Bayesian updating with classical probabilities, where prior and posterior fit nicely together in the sense that

i) they belong to the same class of parametric distributions (*conjugate* prior)

ii) the updating of one parameter of the prior is linear.

Definition (LUCK-model)

- classical Bayesian inference on ϑ based on sample w as in (1)
- prior $p(\vartheta)$ characterized by (vectorial) parameter $\vartheta^{(0)}$.

Call $\big(p(\vartheta), p(\vartheta \mid w)\big)$ LUCK-*model of size q in natural parameter* ψ *with prior param.s* $n^{(0)} \in \mathbb{R}^+$ *and* $y^{(0)}$ *and sample statistic* $\tau(w)$

$\iff$

$\exists\ q \in \mathbb{N}$, transformations $\vartheta \mapsto \psi$, $\vartheta \mapsto \mathbf{b}(\psi)$, $\vartheta^{(0)} \mapsto (n^{(0)}, y^{(0)})$ such that

$$p(\vartheta) \propto \exp\left\{n^{(0)}\left[\langle\psi, y^{(0)}\rangle - \mathbf{b}(\psi)\right]\right\} \tag{2}$$

$$\text{and} \qquad p(\vartheta \mid w) \propto \exp\left\{n^{(1)}\left[\langle\psi, y^{(1)}\rangle - \mathbf{b}(\psi)\right]\right\}, \quad \text{where} \tag{3}$$

$$n^{(1)} = n^{(0)} + q \qquad \text{and} \qquad y^{(1)} = \frac{n^{(0)} y^{(0)} + \tau(w)}{n^{(0)} + q}. \tag{4}$$

Sets of Conjugate Priors

- Linear Regression Analysis
  - linear regression & generalizations
  - estimation methods
- Conjugate Priors
  - Bayesian estimation
  - LUCK-models
- Sets of Conjugate Priors
  - method of Quaeghebeur and de Cooman
    - direct application
    - generalizing the standard approach
  - the *imprecise normal regression model*
- Application and Concluding Remarks

Method of Quaeghebeur and de Cooman

- Quaeghebeur and de Cooman (2005) developed a method to create sets of conjugate priors for exponential family sample distributions that is easy to handle

> central idea: formulate prior not in classical parameters, but in so-called natural parameters $y^{(0)}$ and $n^{(0)}$ in (2) and (3)
> ▶ update step is linear ▶ inf $\longrightarrow$ inf and sup $\longrightarrow$ sup

- very general and powerful model, as exponential family includes most of every-day distributions
- IDM contained as special case of multinomial sampling model with conjugate Dirichlet priors ($y^{(0)} \leftrightarrow t$, $n^{(0)} \leftrightarrow s$)

Extension of the Method of Quaeghebeur and de Cooman

note: argument is

- not limited to i.i.d. samples
- not limited to the way of contruction of priors

Extension of the Method of Quaeghebeur and de Cooman

note: argument is

- not limited to i.i.d. samples
- not limited to the way of contruction of priors

idea: use method for LUCK-models, and apply to regression

Procedure of Quaeghebeur and de Cooman

- define set of priors via set of parameters
- define this set of parameters by lower and upper bounds
- lower and upper bounds of set of posterior parameters can be obtained directly from the update formula:

$$y^{(1)} = \frac{n^{(0)} y^{(0)} + \tau(w)}{n^{(0)} + q}$$

Just as in the IDM, minimization and maximization problems on the set of posteriors are reduced to minimization and maximization problems on the set of priors when parameter $y^{(1)}$ (or a linear function of it) is the quantity of interest.

Conjugate Priors for Linear Regression

(at least) two possibilities for LUCK-model for linear regression:

1. construct prior to likelihood as described in Quaeghebeur and de Cooman (2005) / Bernardo and Smith (1993)
   ▶ $\mathbf{X}$ is part of prior; can be shown to be normal at least for $p = 2$ ▶ approach only sketched
2. take well-known standard conjugate prior
   ▶ fits to method of Quaeghebeur and de Cooman, as it can be shown to constitue a LUCK-model for arbitrary number p of regressors (Theorem 2)

2. Standard Conjugate Prior

called *normal regression model* in the paper

$$\beta \sim \mathrm{N}_p\left(\beta^{(0)}, \sigma^2 \boldsymbol{\Sigma}^{(0)}\right)$$

$$\beta \,|\, z \sim \mathrm{N}_p\left(\beta^{(1)}, \sigma^2 \boldsymbol{\Sigma}^{(1)}\right) ,$$

where the updated parameters $\beta^{(1)}$ and $\boldsymbol{\Sigma}^{(1)}$ are obtained as

$$\beta^{(1)} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X} + \boldsymbol{\Lambda}^{(0)}\right)^{-1}\left(\mathbf{X}^{\mathsf{T}} z + \boldsymbol{\Lambda}^{(0)}\beta^{(0)}\right)$$

$$\boldsymbol{\Sigma}^{(1)} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X} + \boldsymbol{\Lambda}^{(0)}\right)^{-1} ,$$

$\boldsymbol{\Lambda}^{(0)} = {\boldsymbol{\Sigma}^{(0)}}^{-1}$ being the so-called *precision matrix*.

The *Imprecise Normal Regression Model*

Theorem

Fixing a value $n^{(0)}$, $\big(p(\beta), p(\beta \mid z)\big)$ *constitutes a* LUCK*-model of size* 1 *with prior parameters*

$$y^{(0)} = \frac{1}{n^{(0)}} \begin{pmatrix} \mathbf{\Lambda}^{(0)} \\ \mathbf{\Lambda}^{(0)} \beta^{(0)} \end{pmatrix} =: \begin{pmatrix} y_a^{(0)} \\ y_b^{(0)} \end{pmatrix}$$

and $n^{(0)}$ *and sample statistic*

$$\tau(z) = \tau(\mathbf{X}, z) = \begin{pmatrix} \mathbf{X}^{\mathsf{T}}\mathbf{X} \\ \mathbf{X}^{\mathsf{T}} z \end{pmatrix} =: \begin{pmatrix} \tau_a(\mathbf{X}, z) \\ \tau_b(\mathbf{X}, z) \end{pmatrix}.$$

Proof: The proof is given in Walter (2006) and Walter (2007B)

'Translation' Issues

1. Express prior knowledge on β by a set of $\beta^{(0)}$'s and $\mathbf{\Lambda}^{(0)}$'s.
2. "Translate" this set into set of $y^{(0)}$'s such that resulting set $\mathcal{Y}^{(0)}$ consists only of admissible combinations of parameters (positive definiteness of $\mathbf{\Lambda}^{(0)}$, bounding of $\mathcal{Y}^{(0)}$ as advocated by Quaeghebeur and de Cooman)
3. Update each $y^{(0)}$ in $\mathcal{Y}^{(0)}$ by (4) linearly to $y^{(1)}$.
4. "Retranslate" set $\mathcal{Y}^{(1)}$ into an interpretable set of values of $\beta^{(1)}$ and $\mathbf{\Lambda}^{(1)}$.

2\. highly komplex for arbitrary p

- analytical results derived for $p = 2$ (with further simplifications).
- properties of resulting model very plausible.

Application to Data from the AIRGENE Study

AIRGENE: EU financed panel study

air pollutants $\stackrel{?}{\longrightarrow}$ inflammation markers in myocardial infarction survivors

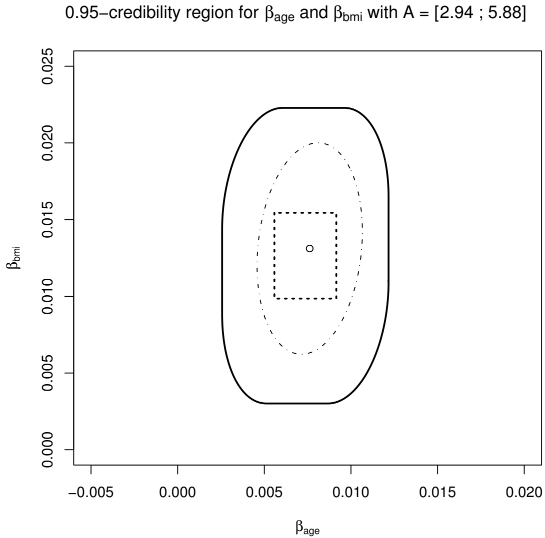
but:

inflammation markers $\longleftrightarrow$ BMI (Body-Mass-Index) and age

- must be taken into account to adjust air pollutants $\longrightarrow$ inflammation markers.

Model:

$$\log(\texttt{fib})_i = [\underline{\beta}_0, \overline{\beta}_0] + \texttt{age}_i \cdot [\underline{\beta}_{\texttt{age}}, \overline{\beta}_{\texttt{age}}] + \texttt{bmi}_i \cdot [\underline{\beta}_{\texttt{bmi}}, \overline{\beta}_{\texttt{bmi}}] + \varepsilon_i$$



Very low 'trust' in prior information corresponding to 1 – 2 observations

Concluding Remarks: Overview

- Linear Regression Analysis
  - linear regression & generalizations
  - estimation methods
- Conjugate Priors
  - Bayesian estimation
  - LUCK-models
- Sets of Conjugate Priors
  - method of Quaeghebeur and de Cooman
    - direct application
    - generalizing the standard approach
  - the *imprecise normal regression model*
- Application and Concluding Remarks

Concluding Remarks: Method of QdC and Prior Data Conflict

Quaeghebeur and de Cooman vary $y^{(0)}$ in a set and fix $n^{(0)}$
(IDM: vary $t_1, \ldots, t_k$, fix s)

- insufficient behavior in case of prior-data conflict, as

$$\overline{y}^{(1)} - \underline{y}^{(1)} = \frac{n^{(0)} \left(\overline{y}^{(0)} - \underline{y}^{(0)}\right)}{n^{(0)} + n}$$

when $y^{(0)}$ varies between $\underline{y}^{(0)}$ and $\overline{y}^{(0)}$

- imprecision decreases by same amount for any sample of size n

possible solution: vary $n^{(0)}$ in addition: to be explored in generality, but already done by Walley (1991, Ch. 5.4) for two-parameter IDM.

- updating of $y^{(0)}$ non-linear!